Lab 4

Levon Demirdjian

Tuesday, April 19, 2016

Problem 1
A random variable X has the following cdf:

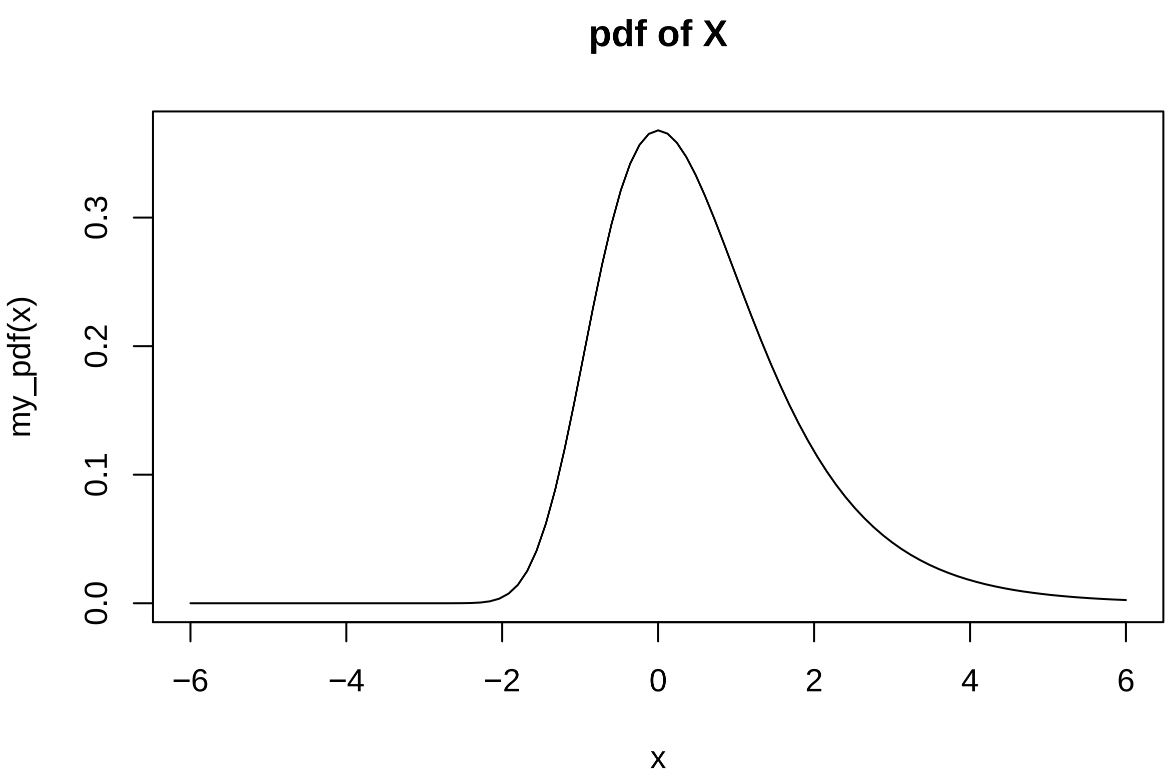
$$F(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}.$$

We may be interested in computing the mean and variance of this distribution, but may find it difficult to find an exact expression for these quantities. We can use Monte Carlo integration to approximate these quantities instead.

1. Find the pdf of X, $f_X(x)$.
2. Plot the pdf to get a sense of which points are more probable.
3. Sample from $f_X(x)$ using the inversion method.

The pdf is $f_X(x) = e^{-x}e^{-e^{-x}}$.

```
my_pdf <- function(x){exp(-x)*exp(-exp(-x))}
curve(my_pdf, -6, 6, main = "pdf of X")
```



To sample from $f_X(x)$, we must first find the inverse cdf:

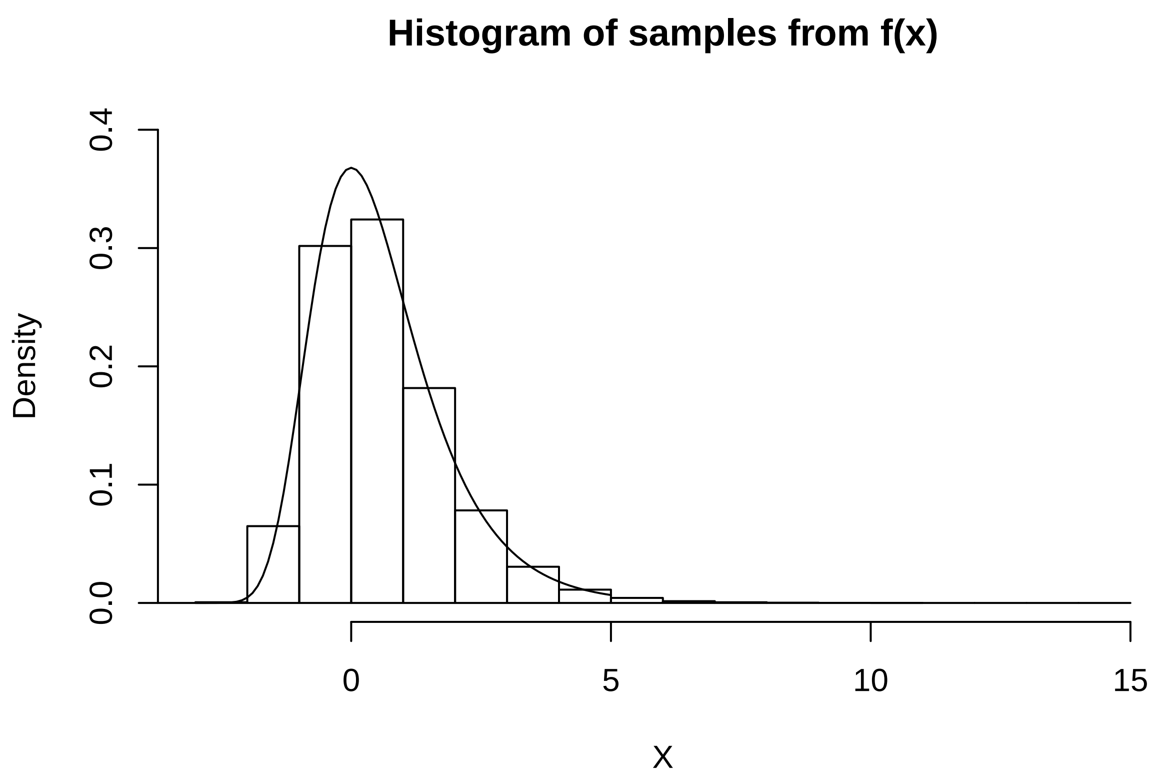
$$F(x) = u \Leftrightarrow e^{-e^{-x}} = u \Leftrightarrow -e^{-x} = \log(u) \Leftrightarrow x = -\log(-\log(u)).$$

Therefore,

$$F^{-1}(u) = -\log(-\log(u)).$$

Now we can generate a sample from f_X and plot our results:

```
my_invcdf <- function(x){-log(-log(x))}
U <- runif(1000000)
X <- my_invcdf(U)
hist(X, prob = TRUE, main = "Histogram of samples from f(x)", ylim = c(0, 0.4))
curve(my_pdf, -5, 5, add = TRUE)
```



4. Say we want to find $E(X)$. Write the expression for $E(X)$; can you find this analytically? What about $Var(X)$?
5. Design a method to approximate the above quantities.

Finding exact expressions for $E(X)$ and $Var(X)$ is difficult. It turns out that the mean of this distribution is given by Euler's constant

$$\gamma = \lim_{n\to\infty}\left(-\log(n) + \sum_{k=1}^{n}\frac{1}{k}\right) \approx 0.577,$$

and the variance is $\frac{\pi^2}{6}$.

We can proceed via Monte Carlo integration to find an approximation of the mean and variance. Recall that if $X_1, ..., X_N \overset{iid}{\sim} f_X(x)$, then by the law of large numbers,

$$\frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{p} E(X),$$

i.e. the sample average of points sampled from the distribution approaches its expectation as $N \to \infty$. We can use this fact to approximate the mean:

```
mean(X)
```

```
## [1] 0.5790474
```

```
-digamma(1) ## This is Euler's constant
```

```
## [1] 0.5772157
```

What about the variance? Since $Var(X) = E[(X - E(X))^2]$, we can use the LLN again:

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 \xrightarrow{p} Var(X).$$

Note that strictly speaking, $E[\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2] = \frac{N-1}{N}Var(X)$, but in the limit this doesn't matter.

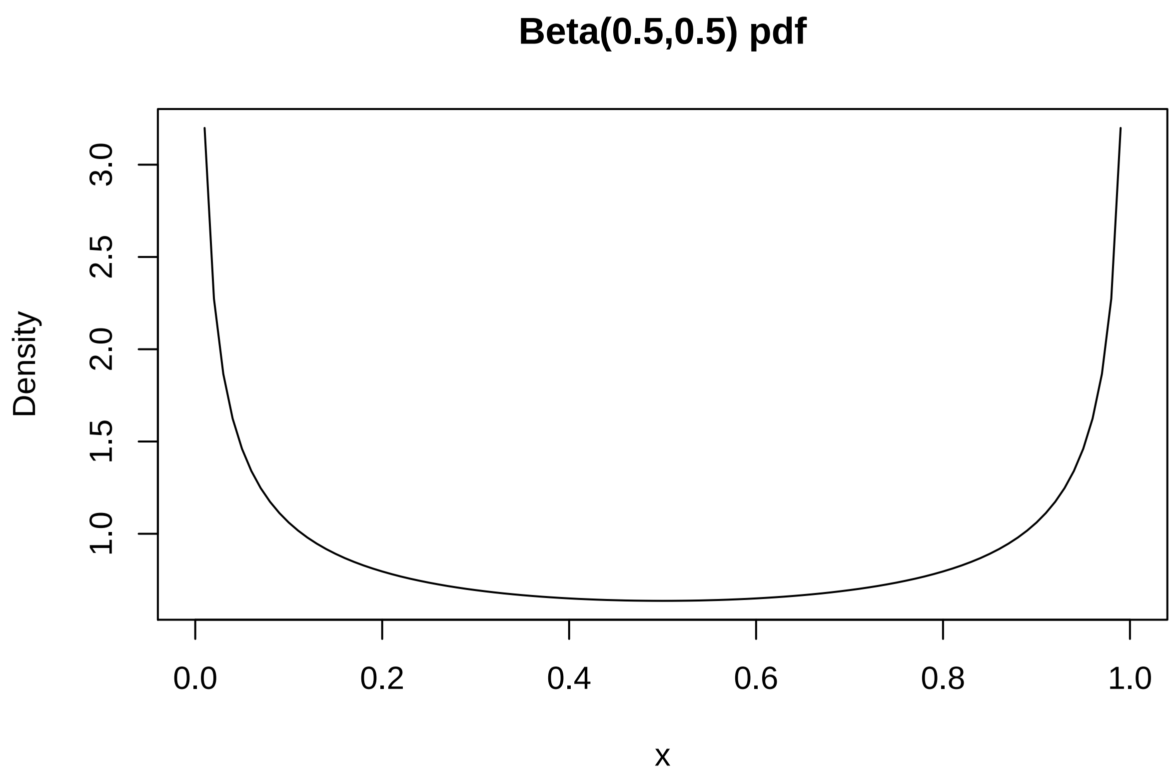
```
mean((X-mean(X))^2)
```

```
## [1] 1.642296
```

```
(pi^2)/6
```

```
## [1] 1.644934
```

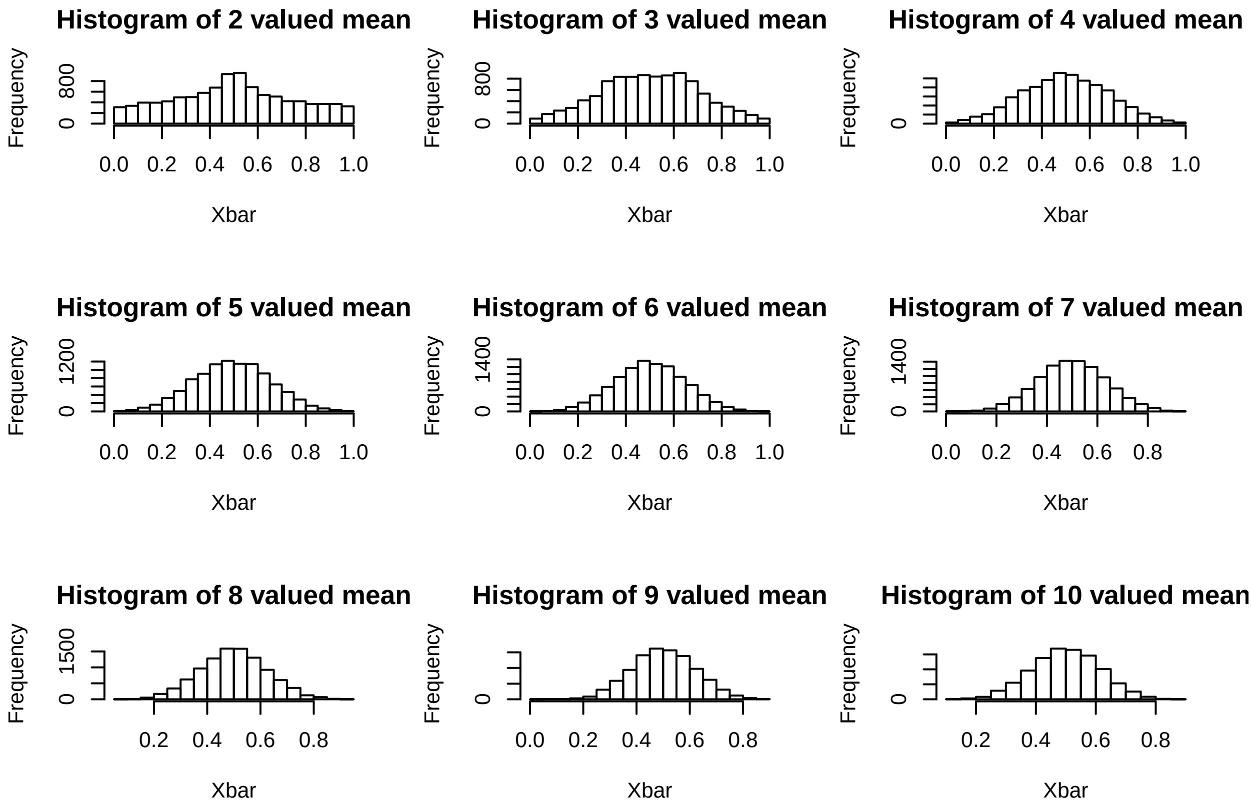
Problem 2
Consider the following beta(0.5, 0.5) distribution: $f(x) = cx^{-0.5}(1-x)^{-0.5}$ where $c > 0$ is the normalizing constant, and $0 \le x \le 1$. The pdf looks like this:



1. Sample X_1, X_2 iid from $f(x)$ using the rbeta function in R, and compute their mean. Do this a large number of times (say 1000 times) and plot a histogram of the means.
2. Repeat the above procedure for samples of size $n = 3, 4, ..., 10$.
3. What do you notice?

```
results <- matrix(0, nrow = 10, ncol = 10000)
for(n in 1:10){
  for(i in 1:10000){
    results[n, i] <- mean(rbeta(n, 0.5, 0.5))
  }
}

par(mfrow = c(3, 3))
for(i in 2:10){
  hist(results[i,], main = paste("Histogram of", i, "valued mean"), xlab = "Xbar")
}
```


Histogram of 2 valued mean
Frequency
0 800
0.0 0.2 0.4 0.6 0.8 1.0
Xbar
Histogram of 3 valued mean
Frequency
0 800
0.0 0.2 0.4 0.6 0.8 1.0
Xbar
Histogram of 4 valued mean
Frequency
0
0.0 0.2 0.4 0.6 0.8 1.0
Xbar
Histogram of 5 valued mean
Frequency
0 1200
0.0 0.2 0.4 0.6 0.8 1.0
Xbar
Histogram of 6 valued mean
Frequency
0 1400
0.0 0.2 0.4 0.6 0.8 1.0
Xbar
Histogram of 7 valued mean
Frequency
0 1400
0.0 0.2 0.4 0.6 0.8
Xbar
Histogram of 8 valued mean
Frequency
0 1500
0.2 0.4 0.6 0.8
Xbar
Histogram of 9 valued mean
Frequency
0
0.0 0.2 0.4 0.6 0.8
Xbar
Histogram of 10 valued mean
Frequency
0
0.2 0.4 0.6 0.8
Xbar

−>